to a general medical ward with left hip pain, persisting anaemia, and confusion. Examination revealed a large abscess of his left hip and radiographs showed almost complete destruction of the femoral head. Blood cultures grew *S aureus*. His C reactive protein peaked at 122 mg/l. The hip was drained surgically of large amounts of pus, culture of which grew *S aureus* and *Proteus mirabilis*. He was treated with a prolonged course of high dose antibiotics, with some clinical improvement but continuing poor mobility.

Discussion

These patients were all elderly and had pre-existing osteoarthritis and concurrent infection elsewhere. None, however, had other systemic conditions predisposing to infection, such as diabetes, except for the second patient, who had a myeloproliferative disorder. The development of septic arthritis by haematogenous spread was associated with increasing hip pain and rapid destruction of the femoral head. This was accompanied by a delay in diagnosis of up to six months.

Infection in the presence of existing inflammatory joint disease, particularly rheumatoid arthritis, is well known.[1] It is much rarer to see this in association with the much commoner osteoarthritis, although it is recognised.[2] In common with other bone and joint infections, the presentation of septic arthritis has changed in recent years from the usual florid illness.

The clinical features may be muted, particularly in the elderly[3] and especially when attenuated by a course of antibiotics given to treat the original concurrent infection. If the diagnosis is missed or delayed, the consequences are serious: the joint destruction may preclude successful arthroplasty or, perhaps worse, a hip replacement may be inserted into an unrecognised septic environment.

The most useful non-specific tests seem to be the erythrocyte sedimentation rate and measurement of C reactive protein; the single most useful specific test is joint aspiration and culture.

We recommend consideration of septic arthritis in any patient with an apparently acute exacerbation of an osteoarthritic joint, particularly if there is a possibility of coexistent infection elsewhere. Other possible non-infective causes of a rapid deterioration in symptoms include pseudogout and avascular necrosis, and these will also need to be considered.

1 Gardner GC, Weisman MH. Pyarthrosis in patients with rheumatoid arthritis: a report of 13 cases and a review of the literature from the past 40 years. *Am J Med* 1990;88:503-11.
2 Goldenberg DL, Cohen AS. Acute infectious arthritis. *Am J Med* 1976;60:369-77.
3 Vincent GM, Amirault JD. Septic arthritis in the elderly. *Clin Orthop* 1991;251:241-5.

Statistics Notes

Measurement error and correlation coefficients

J Martin Bland, Douglas G Altman

This is the 22nd in a series of occasional notes on medical statistics

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *professor of medical statistics*

ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, PO Box 777, Oxford OX3 7LF
Douglas G Altman, *head*

Correspondence to:
Professor Bland.

BMJ 1996;313:41-2

Measurement error is the variation between measurements of the same quantity on the same individual.[1] To quantify measurement error we need repeated measurements on several subjects. We have discussed the within-subject standard deviation as an index of measurement error,[1] which we like as it has a simple clinical interpretation. Here we consider the use of correlation coefficients to quantify measurement error.

A common design for the investigation of measurement error is to take pairs of measurements on a group of subjects, as in table 1. When we have pairs of observations it is natural to plot one measurement against the other. The resulting scatter diagram (see figure 1) may tempt us to calculate a correlation coefficient between the first and second measurement. There are difficulties in interpreting this correlation coefficient. In general, the correlation between repeated measurements will depend on the variability between subjects. Samples containing subjects who differ greatly will produce larger correlation coefficients than will samples containing similar subjects. For example, suppose we split this group in whom we have measured forced expiratory volume in one second (FEV_1) into two subsamples, the first 10 subjects and the second 10 subjects. As table 1 is ordered by the first FEV_1 measurement, both subsamples vary less than does the whole sample. The correlation for the first subsample is r = 0.63 and for the second it is r = 0.31 , both less than r = 0.77 for the full sample. The correlation coefficient thus depends on the way the sample is chosen, and it has meaning only for the population from which the study subjects can be regarded as a random sample. If we select subjects to give a wide range of the measurement, the

natural approach when investigating measurement error, this will inflate the correlation coefficient.

The correlation coefficient between repeated measurements is often called the reliability of the measurement method. It is widely used in the validation of psychological measures such as scales of anxiety and depression, where it is known as the test-retest reliability. In such studies it is quoted for different populations (university students, psychiatric outpatients, etc) because the correlation coefficient differs between them as a result of differing ranges of the quantity being measured. The user has to select the correlation from the study population most like the user's own.

Another problem with the use of the correlation coefficient between the first and second measurements is

Table 1—*Pairs of measurements of FEV_1 (litres) a few weeks apart from 20 Scottish schoolchildren, taken from a larger study (D Strachan, personal communication)*

Subject No	Measurement 1st	Measurement 2nd	Subject No	Measurement 1st	Measurement 2nd
1	1.19	1.37	11	1.54	1.57
2	1.33	1.32	12	1.59	1.60
3	1.35	1.40	13	1.61	1.53
4	1.36	1.25	14	1.61	1.61
5	1.38	1.29	15	1.62	1.68
6	1.38	1.37	16	1.78	1.76
7	1.38	1.40	17	1.80	1.82
8	1.40	1.38	18	1.85	1.89
9	1.43	1.38	19	1.94	2.10
10	1.43	1.51	20	2.10	2.20

Table 2—*One way analysis of variance for the data in table 1*

Source of variation	Degrees of freedom	Sum of squares	Mean square	Variance ratio (F)	Probability (P)
Children	19	1.52981	0.08052	7.4	<0.0001
Residual	20	0.21670	0.01086		
Total	39	1.74651			

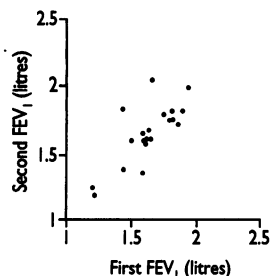**Fig 1**—*Measurements from pairs of observations plotted against each other*

that there is no reason to suppose that their order is important. If the order were important the measurements would not be repeated observations of the same thing. We could reverse the order of any of the pairs and get a slightly different value of the correlation coefficient between repeated measurements. For example, reversing the order of the even numbered subjects in table 1 gives r = 0.80 instead of r = 0.77. The intra-class correlation coefficient avoids this problem. It estimates the average correlation among all possible orderings of pairs. It also extends easily to the case of more than two observations per subject, where it estimates the average correlation between all possible pairs of observations.

Few computer programs will calculate the intra-class correlation coefficient directly, but when the number of observations is the same for each subject it can be found from a one way analysis of variance table[2] such as table 2. We need the total sum of squares, SS_T, and the sum of squares between subjects, SS_B.

Then

$$r_I = \frac{mSS_B - SS_T}{(m - 1) \, SS_T}$$

where m is the number of observations per subject. For table II, $m = 2$ and

$$r_I = \frac{2 \times 1.52981 - 1.74651}{(2 - 1) \times 1.74651} = 0.75$$

In practice, there will usually be little difference between r and r_I for true repeated measurements. If, however, there is a systematic change from the first measurement to the second, as might be caused by a learning effect, r_I will be much less than r. If there was such an effect the measurements would not be made under the same conditions and so we could not measure reliability.

The correlation coefficient can be used to compare measurements of different quantities, such as different scales for measuring anxiety. We could make repeated measurements of all the quantities on the same subjects and calculate intra-class correlations. The measures with the highest correlation between repeated measurements would discriminate best between individuals; in other words they would carry the most information. For most applications, however, we prefer the within-subjects standard deviation as an index of measurement error, as it has a more direct interpretation which can be applied to individual measurements.[1]

1 Bland JM, Altman DG. Measurement error. *BMJ* 1996;312:1654.
2 Altman DG, Bland JM. Comparing several groups using a analysis of variance *BMJ* 1996;312:1472-3.

Correction

Grand Rounds—Hammersmith Hospital: A physiology classic revisited

Two editorial errors occurred in this article by A Al-Mohammad (20 April, pp 1029-31). In the subtitle and throughout the text "one [or two] kidney, one clipped model [or hypertension]" should have read "one [or two] kidney, one clip model [or hypertension]." In figure 2 the angiograms were reversed—the left hand angiogram shows stenosis after angioplasty, and the right hand one shows stenosis before angioplasty.

A MEMORABLE PATIENT

A fishy tale

I never liked surgery. A few fumbling appendicectomies as a house surgeon cured me of any surgical ambitions, and from the day that I started my anaesthetic training I contented myself with telling surgeons what to do. But occasionally, you have to rise to the occasion.

One day in the late 1960s my children told me that one of their pet goldfish was standing on its head. Inspection proved that this was no exaggeration. One of the goldfish had adopted a vertical posture, nose down, on the bottom of the tank. I gently disturbed the fish and it swam around, apparently quite happy, but as soon as it stopped swimming it slowly sank down, nose first, until it was, once again, standing on its head. After this had happened two or three times I gave it a closer inspection and saw that a small pebble had become wedged in its mouth. Clearly this had upset its buoyancy and only by active swimming could it maintain a horizontal posture.

Seeking the advice of a veterinary surgeon for a fish that insisted on standing on its head would I thought invite scepticism, if not ridicule, so I decided to tackle the problem myself.

Soaking a cloth in the water, I wrapped it round the fish and using a hypodermic needle tried to dislodge the pebble—all to no avail. Encouraged by my young audience, I overcame my dislike of surgery and used the needle to make a small incision under the jaw, through which I was then successful in pushing the pebble. I unwrapped the fish, which swam off, apparently none the worse for its operation—or so I thought. As soon as it stopped swimming, however, the fish floated to the surface until its dorsal fin was poking out of the water. Presumably, it had used some adaptive mechanism to try to overcome the change in its buoyancy caused by the pebble. Once the pebble had been removed the adaptive mechanism was overcompensating. Happily, this lasted only a matter of hours, after which it seemed to sort out its buoyancy problems and lived happily ever after.

All this goes to show that even the most inexperienced surgeon may occasionally have a successful outcome, although I have never submitted my results to surgical audit.—JOHN S M ZORAB *is a retired consultant anaesthetist in Bristol*

We welcome filler articles of up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk.